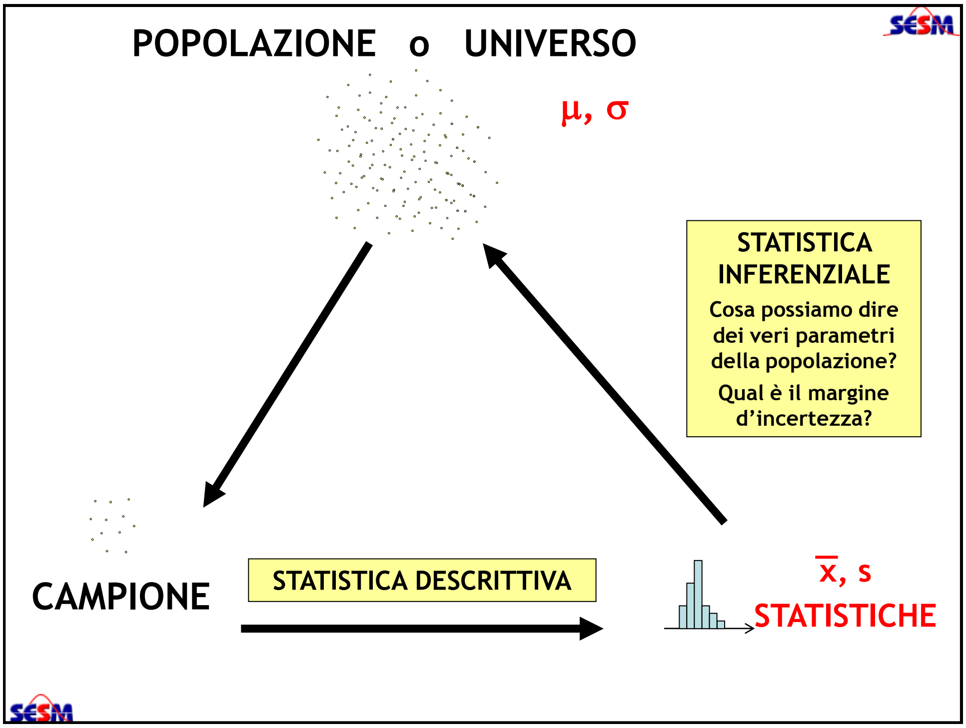
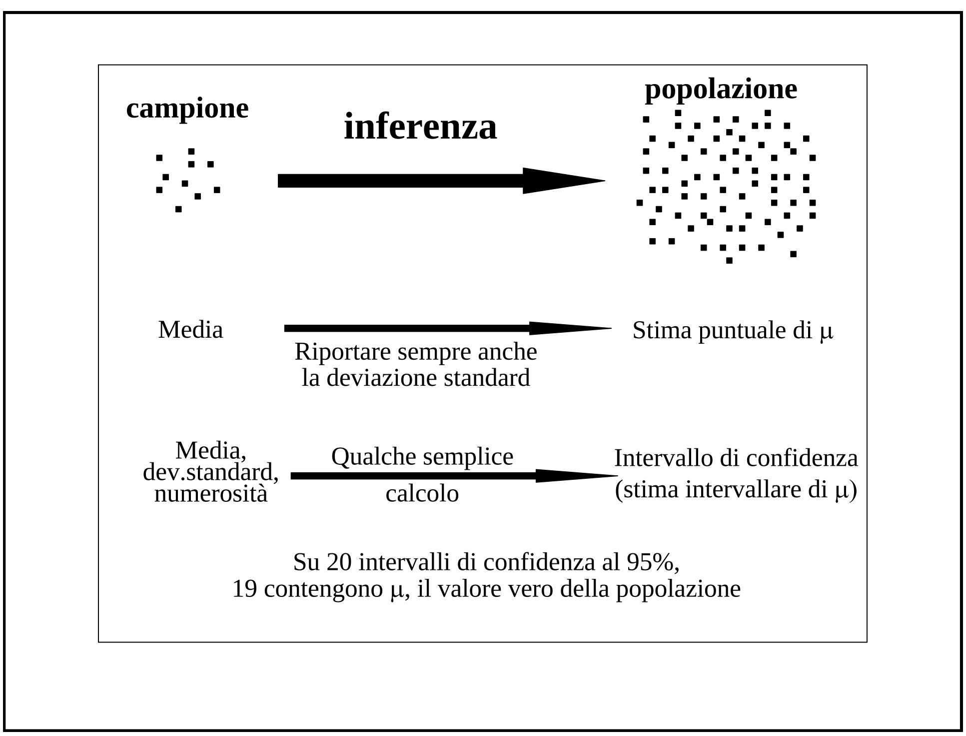
Brevi cenni all'intervallo di confidenza

INFERENZA STATISTICA

L'INFERENZA STATISTICA è un insieme di metodi con cui si cerca di «raggiungere una conclusione» sulla popolazione, sulla base delle informazioni contenute in un campione, estratto da quella popolazione


POPOLAZIONE o UNIVERSO
μ, σ
STATISTICA INFERENZIALE
Cosa possiamo dire dei veri parametri della popolazione?
Qual è il margine d'incertezza?
CAMPIONE
STATISTICA DESCRITTIVA
x̄, s
STATISTICHE
SESM
SESM


campione
inferenza
popolazione
Media
Riportare sempre anche la deviazione standard
Stima puntuale di μ
Media, dev.standard, numerosità
Qualche semplice calcolo
Intervallo di confidenza (stima intervallare di μ)
Su 20 intervalli di confidenza al 95%, 19 contengono μ, il valore vero della popolazione

Dal momento che il campione viene estratto casualmente dalla popolazione, le conclusioni tratte da un campione possono essere errate.

L'inferenza statistica viene fatta "con umiltà":

1) si cerca di stimare la probabilità di commettere errori

2) si cerca di limitare la probabilità di commettere errori

INTERVALLO di CONFIDENZA

Lo scopo dell'inferenza statistica è la conoscenza dei **parametri** che caratterizzano una popolazione.

Per conoscere il parametro, però, dovremmo prendere in esame **tutte** le unità statistiche che costituiscono la popolazione; questo spesso è impossibile perché:

1. numerosità molto elevata
2. spesso la popolazione obiettivo è infinita

impossibile conoscere il **parametro**

Non potendo calcolare con esattezza il parametro, **ricorriamo ad una sua stima**.

La **statistica** (es. $\bar{x}$ s; p) calcolata su un campione estratto dalla popolazione obiettivo è una **stima puntuale** del parametro della popolazione.

Questa stima puntuale del parametro non sarà mai identica al vero parametro della popolazione, ma sarà affetta da un **errore** per eccesso o per difetto.

In molte situazioni è preferibile **una stima intervallare** (cioè è preferibile indicare come stima del parametro un intervallo al posto di un *singolo punto* sull'asse dei valori) che esprima anche l'**errore associato alla stima** (precisione).

Una **STIMA PUNTUALE** è il risultato di un procedimento che, attraverso le informazioni tratte dal campione osservato, genera un singolo valore numerico usato per stimare il corrispondente parametro della popolazione

Una **STIMA INTERVALLARE** è il risultato di un procedimento che, attraverso le informazione tratte dal campione osservato, genera un intervallo di valori che, con un dato grado di fiducia, conterrà il parametro da stimare.

INTERVALLO di CONFIDENZA: DEFINIZIONE

Per intervallo di confidenza di un parametro Θ (ad es. della media μ o della proporzione/prevalenza) della popolazione, intendiamo un intervallo delimitato da due limiti L_{inf} (limite inferiore) ed L_{sup} (limite superiore) che abbia una definita probabilità (1- α) (ad es. (1-0.05)=0.95) di contenere il vero parametro della popolazione:

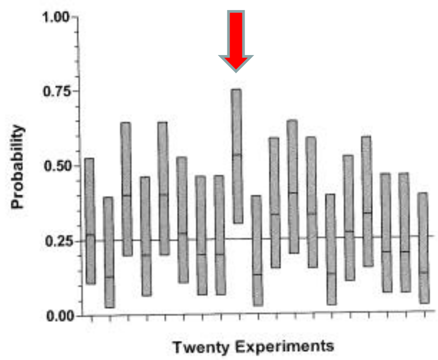
$$p(L_{inf} < \Theta < L_{sup}) = 1-\alpha$$

$$p(L_{inf} < \mu \text{ (o } \pi) < L_{sup}) = 1-0.05 = 0.95$$

dove: $1- \alpha$ = grado di confidenza

α = probabilità di errore

Facciamo questa simulazione partendo da una popolazione di cui conosciamo tutto: da un'urna con **100** palline, **25 rosse** e **75 nere**, scegliamo n=15 palline, calcoliamo la proporzione di rosse e il relativo intervallo di confidenza al **95% (IC(95%)).** Ripetiamo questo processo (esperimento) per 20 volte



Esempio con IC della proporzione

- Vogliamo stimare la prevalenza del dolore (presenza) negli ospedali italiani:
- prendiamo un campione di 20 ospedali e abbiamo il dato sulla presenza del dolore di 3575 pazienti: risulta pari a 91,2% *(stima puntuale).*
- Calcoliamo l'intervallo di confidenza al 95% che risulta: 90,3%-92,1% *(stima intervallare)*

questo intervallo ha una probabilità del 95% di contenere la vera prevalenza del dolore della popolazione ospedaliera italiana

INTERVALLO di CONFIDENZA di una PROPORZIONE

Per N > 30:

$$p \sim N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

π sarà stimato da p

E che:

$$p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$$

Esempio con IC della media

- Vogliamo stimare il livello medio di glicemia nei diabetici italiani:
- prendiamo un campione di 36 soggetti; la media della glicemia in questo gruppo risulta 155 mg/dl *(stima puntuale).*
- Calcoliamo l'intervallo di confidenza al 95% che risulta: 147,2-162,8 mg/dl *(stima intervallare)*

questo intervallo ha una probabilità del 95% di contenere la vera media della popolazione dei diabetici italiani

Esempio - *continua*

- L'intervallo di confidenza al 95% del lucido precedente è stato ottenuto nel seguente modo:

$$\bar{x} = 155 mg/dl$$

$$s = 24 mg/dl$$

$$n = 36$$

$$\bar{x} \pm 1{,}96 * \frac{s}{\sqrt{n}} = 155 \pm 1{,}96 * \frac{24}{\sqrt{36}}$$

L_{inf}=147,2

L_{sup}=162,8

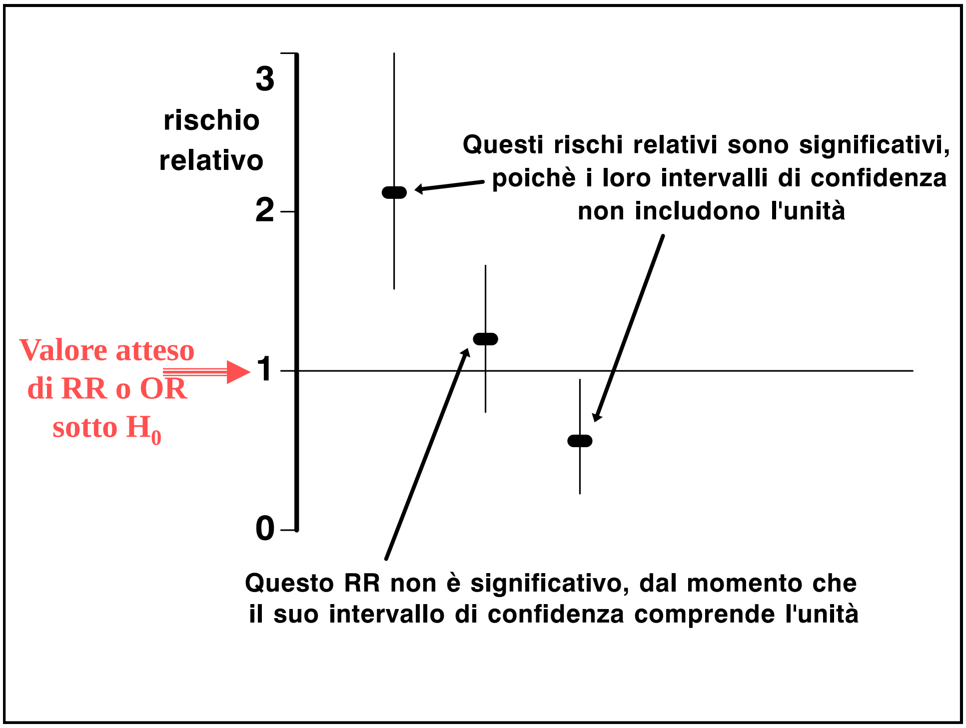
RIASSUMENDO...

La **stima puntuale** fornisce un <u>singolo valore</u>. Tuttavia:

1. questo valore non coincide quasi mai con il valore vero (parametro) della popolazione;
2. campioni diversi forniscono stime puntuali diverse.

La **stima intervallare** fornisce <u>un intervallo</u>:

1. quest'intervallo ha una determinata probabilità (in genere, il 95%) di contenere il valore vero (parametro) della popolazione;


3
rischio
relativo
2
1
0
Questi rischi relativi sono significativi, poichè i loro intervalli di confidenza non includono l'unità
Valore atteso di RR o OR sotto H_0
Questo RR non è significativo, dal momento che il suo intervallo di confidenza comprende l'unità